T S1/5/1

1/5/1
DIALOG(R)File 351:Derwent WPI

012697796 **Image available**
WPI Acc No: 1999-503905/199942
XRPX Acc No: N99-376716
Duplicate data co-ordination management procedure in distributed database system - involves performing strong and weak co-ordination management at different levels depending on variety of transaction occurred in system
Patent Assignee: JISEDAI JOHO HOSO SYSTEM KENYUJO KK (JISE-N); RICOH KK (RICO)
Number of Countries: 001 Number of Patents: 002
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
JP 11219309	A	19990810	JP 98112305	A	19980422	199942	B
JP 3563591	B2	20040908	JP 98112305	A	19980422	200459	

Priority Applications (No Type Date): JP 97328217 A 19971128; JP 97264643 A 19970929
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 11219309	A		21	G06F-012/00	
JP 3563591	B2		29	G06F-012/00	Previous Publ. patent JP 11219309

Abstract (Basic): JP 11219309 A
NOVELTY - Strict co-ordination management is done enabling immediate and delayed updating transmission from main site to other sites respectively for first two levels. Weak co-ordination management is done for third level, delaying updating transmission from main site to other site. DETAILED DESCRIPTION - During transaction, the level of co-ordination control for reading or updating duplicate data is decided based on transaction variety. An INDEPENDENT CLAIMS are also included for the following: duplicate data co-ordination management procedure; recording medium stored with computer readable co-ordination management program
USE - In distributed database system.
ADVANTAGE - Optimize co-ordination control as it is done depending on variety of transaction.
Dwg.1/14
Title Terms: DUPLICATE; DATA; CO; ORDINATE; MANAGEMENT; PROCEDURE; DISTRIBUTE; DATABASE; SYSTEM; PERFORMANCE; STRONG; WEAK; CO; ORDINATE; MANAGEMENT; LEVEL; DEPEND; VARIETY; TRANSACTION; OCCUR; SYSTEM
Derwent Class: T01
International Patent Class (Main): G06F-012/00
International Patent Class (Additional): G06F-015/16; G06F-017/30
File Segment: EPI
?

JPA 11-219309

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平11－219309

(43)公開日　平成11年(1999)８月10日

(51)Int.Cl.[6]	識別記号	Ｆ Ｉ		
G06F 12/00	518	G06F 12/00	518	A
	533		533	J
15/16	370	15/16	370	M
17/30		15/40	310	C
			310	F

審査請求　未請求　請求項の数７　ＯＬ　（全21頁）　最終頁に続く

(21)出願番号　特願平10－112305

(22)出願日　平成10年(1998)４月22日

(31)優先権主張番号　特願平9－264643
(32)優先日　平９(1997)９月29日
(33)優先権主張国　日本（ＪＰ）
(31)優先権主張番号　特願平9－328217
(32)優先日　平９(1997)11月28日
(33)優先権主張国　日本（ＪＰ）

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(71)出願人　597136766
株式会社次世代情報放送システム研究所
東京都台東区西浅草１丁目１－１
(72)発明者　吉浦　由香利
東京都大田区中馬込１丁目３番６号　株式会社リコー内
(72)発明者　飯沢　篤志
東京都大田区中馬込１丁目３番６号　株式会社リコー内
(74)代理人　弁理士　酒井　宏明

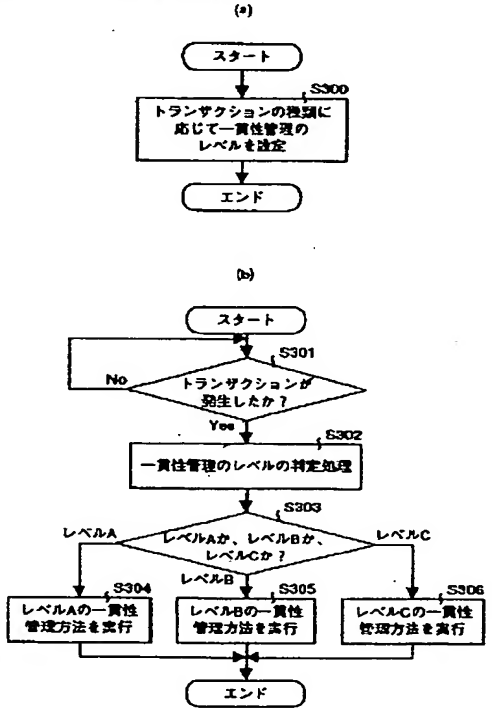
最終頁に続く

(54)【発明の名称】分散型データベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体

(57)【要約】

【課題】　トランザクションの種類に応じて異なる一貫性管理を行うこと。

【解決手段】　各トランザクションの種類に一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定し（Ｓ３００）、システム内においてトランザクションが発生した場合に（Ｓ３０１）、発生したトランザクションの種類に基づいて一貫性管理のレベルを判定し（Ｓ３０２，Ｓ３０３）、第１レベルの場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を即時行い（Ｓ３０４）、第２レベルの場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を遅延させて行い（Ｓ３０５）、第３レベルの場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、主サイトから各サイトへの更新伝播を遅延させて行う（Ｓ３０６）。



1

【特許請求の範囲】

【請求項１】　任意のオブジェクトに対する複製データをそれぞれ有する第１のサイトと、前記複数の第１のサイトにおける複製データの一貫性を集中管理する第２のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、

予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、

システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、

前記判定工程で第１レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、

前記判定工程で第２レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第２の一貫性管理工程と、

前記判定工程で第３レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、

を含むことを特徴とする分散型データベースシステムの一貫性管理方法。

【請求項２】　任意のオブジェクトに対する複製データをそれぞれ有する第１のサイトと、前記複数の第１のサイトにおける複製データの一貫性を集中管理する第２のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、

予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、

システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、

前記判定工程で第１レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、

前記判定工程で第２レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの

2

更新伝播を遅延させて行う第２の一貫性管理工程と、

前記判定工程で第３レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、

を含むことを特徴とする分散型データベースシステムの一貫性管理方法。

【請求項３】　前記第２および第３の一貫性管理工程における更新伝播の遅延とは、第２のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うことであることを特徴とする請求項１または２に記載の分散型データベースシステムの一貫性管理方法。

【請求項４】　前記第２および／または第３の一貫性管理工程における更新伝播の遅延とは、前記第１のサイトにおいて前記トランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第２のサイトにおいて前記更新伝播条件を満足するように複数の第１のサイトへの更新伝播を行うことであることを特徴とする請求項１～３のいずれか一つに記載の分散がデータベースシステムの一貫性管理方法。

【請求項５】　前記第２および／または第３の一貫性管理工程は、少なくとも通信回線および放送波を用いて第２のサイトから複数の第１のサイトへの更新伝播を行うことができ、

前記更新伝播条件は、前記通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むことを特徴とする請求項４に記載の分散型データベースシステムの一貫性管理方法。

【請求項６】　前記判定工程は、前記一貫性管理のレベルを判定することに加えて、データの新規登録か否かを判定し、前記新規登録であると判定した場合に、設定されているレベルに関係なく、前記第３レベルと判定することを特徴とする請求項１～５のいずれか一つに記載の分散型データベースシステムの一貫性管理方法。

【請求項７】　前記請求項１～６のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、分散型データベースシステムにおける複製データの一貫性管理方法に関し、より詳細には、一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができるようにすることにより、分散型データベースシステムの利便性の向上を図った分散型データ

ベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【０００２】

【従来の技術】分散型データベースシステムにおいては、同一データについての複製データが複数のサイトにそれぞれ分散しており、かつ、複数のトランザクションが並行して実行されることになるため、システム上に存在する複製データ間の一貫性をいかに保つかが重要な課題となっている。

【０００３】分散型データベースシステムにおける一貫性管理方法の一例として、データの変更を直列化する操作を行う強い一貫性管理と呼ばれる方法がある。この方法によれば、直列化可能性を保証するが故に操作の並列性を低下させ、一貫性管理のコストが増大するという短所があるものの、直列化可能性が保証されるため、複製データ間の矛盾が一切生じないという長所がある。この強い一貫性管理は、銀行のＯＬＴＰ（Ｏｎ－Ｌｉｎｅ Ｔｒａｎｓａｃｔｉｏｎ Ｐｒｏｃｅｓｓｉｎｇ）等に適用されていることが多い。

【０００４】また、一貫性管理方法の他の例として、複製データ間に矛盾が生じることを認める弱い一貫性管理と呼ばれる方法がある。この弱い一貫性管理では、一つのトランザクションの中でシステム全体の一貫性を保証するのではなく、別のトランザクションで更新伝播を行う。この方法によれば、複製データ間に矛盾が生じてしまうことがあるという短所があるものの、適用する応用によってはその矛盾も問題とならない場合があり、一貫性保持に要するコストを軽減することができるという長所がある。この弱い一貫性管理は、データの更新が少ない安定したシステムに適用されていることが多い。

【０００５】なお、弱い一貫性管理方式については各種の研究がなされており、大別すると、（１）マルチバージョン方式（Ｍｕｌｔｉ－ｖｅｒｓｉｏｎ Ｃｏｎｔｒｏｌ）、（２）差異限定管理方式（Ｂｏｕｎｄｅｄ Ｄｉｖｅｒｇｅｎｃｅ Ｃｏｎｔｒｏｌ）、および（３）意味ベースの方式（Ｓｅｍａｎｔｉｃ－ｂａｓｅｄＣｏｎｔｒｏｌ）に分類することができる。したがって、分散型データベースシステムの規模やデータ内容を考慮して、どのような弱い一貫性管理方式を利用すべきかを決定する必要がある。

【０００６】

【発明が解決しようとする課題】しかしながら、従来の分散型データベースシステムの一貫性管理方法においては、上述した強い一貫性管理および弱い一貫性管理のいずれか一方しか利用することができないため、複製データの一貫性管理において不便な場合があるという問題点があった。この問題点は、分散型データベースシステムの規模・データ内容に応じて、トランザクションの種類毎に強い一貫性管理および弱い一貫性管理を使い分けすることができる方が便利であるという要求に基づくものである。

【０００７】また、弱い一貫性管理を採用した従来の分散型データベースシステムの一貫性管理方法においては、一つのトランザクションの中でシステム全体の一貫性を保証するのではなく、別のトランザクションで更新伝播を行うため、更新伝播がいつ行われるかユーザは認識することができないという問題点があった。すなわち、更新伝播を行うタイミングはシステムに依存しており、ユーザの所望するタイミングで更新伝播を行うことができるようにすることはできなかった。

【０００８】本発明は上記に鑑みてなされたものであって、分散型データベースシステムの一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができるようにすることにより、分散型データベースシステムの利便性の向上を図ることを第１の目的とする。

【０００９】本発明は上記に鑑みてなされたものであって、弱い一貫性管理を採用した場合に、更新伝播を行うタイミングを設定できるようにすることにより、分散型データベースシステムの利便性のさらなる向上を図ることを第２の目的とする。

【００１０】

【課題を解決するための手段】上記目的を達成するため、請求項１の分散型データベースシステムの一貫性管理方法は、任意のオブジェクトに対する複製データをそれぞれ有する第１のサイトと、前記複数の第１のサイトにおける複製データの一貫性を集中管理する第２のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、前記判定工程で第１レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、前記判定工程で第２レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第２の一貫性管理工程と、前記判定工程で第３レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、を含むものである。

【００１１】また、請求項２の分散型データベースシステムの一貫性管理方法は、任意のオブジェクトに対する

複製データをそれぞれ有する第１のサイトと、前記複数の第１のサイトにおける複製データの一貫性を集中管理する第２のサイトと、を備えた分散型データベースシステムの一貫性管理方法において、予め前記複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、前記一貫性管理のレベルを判定する判定工程と、前記判定工程で第１レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、前記判定工程で第２レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第２の一貫性管理工程と、前記判定工程で第３レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、を含むものである。

【００１２】また、請求項３の分散型データベースシステムの一貫性管理方法は、請求項１または２に記載の分散型データベースシステムの一貫性管理方法において、前記第２および第３の一貫性管理工程における更新伝播の遅延について、第２のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うこととしたものである。

【００１３】また、請求項４の分散型データベースシステムの一貫性管理方法は、請求項１～３のいずれか一つに記載の分散がデータベースシステムの一貫性管理方法において、前記第２および／または第３の一貫性管理工程における更新伝播の遅延とは、前記第１のサイトにおいて前記トランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第２のサイトにおいて前記更新伝播条件を満足するように複数の第１のサイトへの更新伝播を行うこととしたものである。

【００１４】また、請求項５の分散型データベースシステムの一貫性管理方法は、請求項４に記載の分散型データベースシステムの一貫性管理方法において、前記第２および／または第３の一貫性管理工程が、少なくとも通信回線および放送波を用いて第２のサイトから複数の第１のサイトへの更新伝播を行うことができ、前記更新伝播条件が、前記通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むものである。

【００１５】また、請求項６の分散型データベースシステムの一貫性管理方法は、請求項１～５のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、前記判定工程が、前記一貫性管理のレベルを判定することに加えて、データの新規登録か否かを判定し、前記新規登録であると判定した場合に、設定されているレベルに関係なく、前記第３レベルと判定するものである。

【００１６】さらに、請求項７のコンピュータ読み取り可能な記録媒体にあっては、前記請求項１～６のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したものである。

【００１７】

【発明の実施の形態】以下、本発明に係る分散型データベースシステムの一貫性管理方法およびその方法の各工程をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体の実施の形態について、添付の図面を参照しつつ詳細に説明する。

【００１８】〔実施の形態１〕実施の形態１に係る分散型データベースシステムの一貫性管理方法においては、複製データを有するサイトが超多地点に存在する分散型データベースシステムに適用することを想定している。

【００１９】ここで、実施の形態１に係る分散型データベースシステムの一貫性管理方法を説明するに先立って、分散型データベースシステムを構築するために考慮すべき二つの条件について説明する。

【００２０】一つ目は、システム内に存在する複製データの更新を行うことができるサイトはどこかということであり、これには集中方式と分散方式とがある。集中方式は、分散型データベースシステムにおける複数のサイトを主サイトと主サイト以外のサイトに分類し、主サイトでのみシステム内に存在する複製データの更新を可能とし、他のサイトではその更新を行うことができないようにするものである。一方、分散方式は、どのサイトでもシステム内の複製データの更新を可能とし（Ａｎｙｓｉｔｅ　Ｗｒｉｔａｂｌｅ）、更新を行うサイトが全ての複製データとの調整を行うというものである。

【００２１】二つ目は、一時的なデータ矛盾を認めるか否かということであり、これには従来技術で説明した強い一貫性管理と弱い一貫性管理とがある。強い一貫性管理は、システム内に存在する複製データ間の矛盾を認めず、一つのトランザクションの中で全ての複製データ間の整合性を保証するものである。例えば、２相コミットプロトコルがこの方式に該当する。一方、弱い一貫性管理は、応用の内容に特化して考えることにより、システム内に存在する複製データ間の矛盾を一時的に認めるものである。この方式では、一つのトランザクションの中

で全ての複製データ間の一貫性が保証されるのではなく、更新の伝播は別のトランザクションで行われる。

【００２２】図１は、集中管理または分散管理と、強い一貫性管理または弱い一貫性管理との組み合わせによって、分散型データベースシステムを構築した場合の長所および短所が変化することを示した説明図である。実施の形態１に係る分散型データベースシステムの一貫性管理方法においては、後述するようにトランザクションの種類に応じて強い一貫性管理および弱い一貫性管理を設定できるようにするために、プロトコルが簡単な集中管理方式を採用することにする。

【００２３】図２は、実施の形態１に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。図２に示す分散型データベースシステムは、複数のサイトからなり、複数のサイトは、任意のオブジェクトから生成した複製データをそれぞれ有するサイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$と、複製データの一貫性を集中管理する主サイトＤＢＳ$_r$と、から構成されている。なお、以下の説明において、サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$をまとめて示す場合にはサイトＤＢＳと記述するものとする。

【００２４】ここで、主サイトＤＢＳ$_r$が管理する情報には、システム内に存在する複製データに対する各オブジェクト毎に、対象オブジェクトＩＤ（いずれのオブジェクトに対する複製データかを特定するためのものであり、各サイトＤＢＳが保有する複製データにも共通のものが付与されている），オブジェクトの最新値，最終更新タイムスタンプ，最終更新サイトＩＤ，最終更新トランザクションＩＤ等がある。一方、各サイトＤＢＳは、システム内に存在する全てのオブジェクトの複製データを有していなければならないというわけではない。さらに、各オブジェクトのマスターデータについては、主サイトＤＢＳ$_r$が集中管理しても良いし（主サイトＤＢＳ$_r$が管理している「オブジェクトの最新値」をマスターデータととらえても良い）、各サイトＤＢＳ毎に管理することにしても良い。

【００２５】なお、図２には、主サイトＤＢＳ$_r$の下にサイトＤＢＳを設けた様子のみが示されているが、各サイトＤＢＳの下にもさらに複数のサイトＤＢＳを設けて階層構造を形成することもできる。このような場合においては、上位のサイトＤＢＳを下位のサイトＤＢＳの主サイトとして動作させることもできる。

【００２６】また、実施の形態１に係る分散型データベースシステムの一貫性管理方法においては、一貫性管理のレベルとしてレベルＡ～Ｃを用意し、トランザクションの種類に応じて、異なるレベルの一貫性管理を行うことができるようにしている。

【００２７】レベルＡ（本発明の第１レベルに該当する）は、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて複製データの一貫性管理を行い、主サイトＤＢＳ$_r$から各サイトＤＢＳへの更新伝播を即時行うというものである。

【００２８】換言すれば、レベルＡは、一つのトランザクションの中で各サイトＤＢＳへの更新伝播まで実行することによって複製データの一貫性管理を行うものであって、トランザクションの直列可能性を保証するものである。

【００２９】また、レベルＢ（本発明の第２レベルに該当する）は、ｒｅａｄ－ｏｎｌｙのトランザクションについては緩和された一貫性管理方法を用いて複製データの一貫性管理を行い、更新を含むトランザクションについては厳格な一貫性管理方法を用いて複製データの一貫性管理を行い、主サイトＤＢＳ$_r$から各サイトＤＢＳへの更新伝播を遅延させて行うというものである。

【００３０】レベルＢにおいては、ある複製データの更新処理を行う場合に、複数のトランザクションに分散して、主サイトＤＢＳ$_r$から各サイトＤＢＳへの更新伝播が行われる。データアクセスに際しては、１コピー直列化可能性は保証する（Ｐ．Ａ．Ｂｅｒｎｓｔｅｉｎ，Ｖ．Ｈａｄｚｉｌａｃｏｓ，ａｎｄ Ｎ．Ｇｏｏｄｍａｎ：Ｃｏｎｃｕｒｒｅｎｃｙ Ｃｏｎｔｒｏｌ ａｎｄ Ｒｅｃｏｖｅｒｙｉｎ Ｄａｔａｂａｓｅ Ｓｙｓｔｅｍｓ，Ａｄｄｉｓｏｎ－Ｗｅｓｌｅｙ，１９８７）。１コピー直列化可能である場合、データ更新の際には同一データに対する複数の複製データの中から最新バージョンを探し、同一データに対する更新処理が競合しないように処理するが、読み出しだけの際には、多少データが古くても良いのでローカルデータをそのまま読むという処理を行う。

【００３１】さらに、レベルＣ（本発明の第３レベルに該当する）は、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、主サイトＤＢＳ$_r$から各サイトＤＢＳへの更新伝播を遅延させて行うというものである。

【００３２】レベルＣは、複製データ間の一時的な一貫性の崩壊を認め、各サイトＤＢＳ毎に複製データの更新を可能とするものである。更新結果は複数のトランザクションとなって主サイトＤＢＳ$_r$から各サイトＤＢＳへ伝播される。この場合、同一データに対する複製データについての更新の競合（衝突）が検出された場合の解決法が問題となる。レベルＣにおいては、後述するように、コミットした後に更新の競合を検出する処理が行われるため、通常の方法のアボートはできない。更新の競合が生じた場合については、後に説明することにする。

【００３３】なお、上述したレベルＡ～Ｃは、後述するように、ｒｅａｄ－ｏｎｌｙまたは更新を含むトランザ

クションの種類に応じて設定される。

【００３４】続いて、実施の形態１に係る分散型データベースシステムの一貫性管理方法を具体的に説明する。図３（ａ）は一貫性管理のレベルを設定する手順を、図３（ｂ）は図３（ａ）に示すようにして設定した一貫性管理のレベルに基づいて一貫性管理方法を実行する手順を示すフローチャートである。ここでは、図３（ａ）および図３（ｂ）を用いて一貫性管理手順の概略を説明した後、レベルＡ～Ｃの一貫性管理方法について具体的に説明する。

【００３５】（一貫性管理手順の概略）まず、図３（ａ）に示すように、分散型データベースシステムの規模・データ内容等を考慮し、システム内で発生するトランザクションの種類に応じて予め一貫性管理のレベルを設定する（Ｓ３００：本発明の一貫性レベル設定工程に該当する）。すなわち、各トランザクションの種類に対して、レベルＡ～Ｃが設定される。なお、一貫性管理のレベルは、分散型データベースシステムのセットアップ時に設定することができ、また、トランザクションの発生時に設定することもできる。

【００３６】図３（ａ）のステップＳ３００においてレベルＡ～Ｃを設定した後、図３（ｂ）に示すように、分散型データベースシステム内にトランザクションが発生すると（Ｓ３０１）、発生したトランザクションに設定されている一貫性管理のレベルの判定処理を実行する（Ｓ３０２：本発明の判定工程に該当する）。

【００３７】そして、ステップＳ３０２における判定処理の結果に基づいて、発生したトランザクションに設定された一貫性管理のレベルがレベルＡ，レベルＢおよびレベルＣのいずれであるかを判定する（Ｓ３０３：本発明の判定工程に該当する）。ステップＳ３０３において、レベルＡであると判定した場合には、レベルＡの一貫性管理方法を実行し（Ｓ３０４：本発明の第１の一貫性管理工程に該当する）、レベルＢであると判定した場合には、レベルＢの一貫性管理方法を実行し（Ｓ３０５：本発明の第２の一貫性管理工程に該当する）、さらに、レベルＣであると判定した場合には、レベルＣの一貫性管理方法を実行する（Ｓ３０６：本発明の第３の一貫性管理工程に該当する）。

【００３８】（レベルＡの一貫性管理方法）次に、レベルＡによる一貫性管理方法について、強い一貫性管理方法の概略について説明した後、実施の形態１のレベルＡによる強い一貫性管理方法を具体的に説明することにする。

【００３９】強い一貫性管理方法としては、２相ロック，楽観的２相ロック，グローバルタイムスタンプに基づく楽観的方法，ベーシックタイムスタンプオーダリング等の各種があり、これらによって一つのトランザクションの中で全ての複製データを更新する場合には直列化可能となる。なお、「完全に複製されたシステムでは、楽観的２相ロックが最も適している」と述べている論文がある（M. J. Carey and M. Livny: “Conflict detectiontradeoffs for replicated data,” ACM TODS, Vol. 16, No. 4, pp. ７０３－７４６, Dec. １９９１：シミュレーションでパフォーマンス評価を行った代表的論文である）。

【００４０】上記強い一貫性管理方法のうち、楽観的制御方式ではコミット時までロックを行わずにプライベート領域（ワークエリア）にのみ複製データの更新結果を保存しておき、コミット時になって初めてトランザクション間の競合があったかどうかを調べる。

【００４１】ここで、あるデータに対する複数の複製データがシステム内に存在する場合の楽観的２相ロック方式の代表例な処理の例を以下に示す（ＩＣＤＣＳ９５のＪｉｎｇのアルゴリズム）。

（１）ｒｅａｄとｗｒｉｔｅに対しては、全複製データＸ０，Ｘ１，・・・，ＸｎをＲｌｏｃｋ（ｒｅａｄモードのロック）する（Ｒｌｏｃｋはトランザクションが複製データを利用していることを示すためのものである）。

（２）トランザクションの終了時に、ｗｒｉｔｅした複製データをＷｌｏｃｋ（ｗｒｉｔｅモードのロック）する。Ｗｌｏｃｋできればトランザクションをコミットし、そうでなければアボートする。

（３）全ロックを解除する。

【００４２】上記楽観的２相ロック方式においては、確認相で全てのサイトＤＢＳにロックをかけなければならないことが問題である。すなわち、実際には全てのサイトＤＢＳにロックをかけることは難しいため、楽観的２相ロック方式は障害に弱いといえる。

【００４３】その理由について、集中管理に基づく２相ロック方式と比較して説明する。集中管理に基づく２相ロック方式の場合は、トランザクションがｒｅａｄしたいとき、マスターデータに対してＲｌｏｃｋをかけてから、ローカルデータを読み出す。ｗｒｉｔｅの場合も同様で、マスターデータにＷｌｏｃｋをかけてから全複製データをｗｒｉｔｅする。したがって、マスターデータのみをロックすれば良いため、負荷を軽減できる。ところが、楽観的２相ロック方式の場合には、全てのサイトでｗｒｉｔｅできなければアボートしなくてはならないため、障害に弱いといえる。

【００４４】また、ダウンしている（参加できない）サイトがあっても、処理が続けられるようにするために、以下の２つの方式が考えられる。

【００４５】第１の方式は、重み付き投票方式（Ｇｉｆｆｏｒｄによって提案されたもの）である。重み付き投票方式においては、以下の条件を満たす一定数のサイトが参加すれば、その中に必ず最新版の値が含まれること

が保証されている。

【００４６】

Ｒｅａｄ－ｑｕｏｒｕｍ+Ｗｒｉｔｅ－ｑｕｏｒｕｍ＞ｎ

Ｗｒｉｔｅ－ｑｕｏｒｕｍ+Ｗｒｉｔｅ－ｑｕｏｒｕｍ＞ｎ

ここで、ｎは複製データの数を表す。

【００４７】第２の方式は、集中管理に基づくタイムスタンプベースの楽観的方式（ＰＴＯ：Ｐｒｉｍａｒｙ－ｂａｓｅｄ　Ｔｉｍｅｓｔａｍｐ　Ｏｐｔｉｍｉｓｔｉｃ Ｃｏｎｔｒｏｌ）である。ＰＴＯ方式は集中管理方式であるため、ダウンしているサイトについては無視することができる。

【００４８】具体的に、集中管理方式では、更新が承認された複製データの最新の値が主サイトＤＢＳ$_P$ に保持される。よって楽観的制御方式の確認相（ｖａｌｉｄａｔｉｏｎ　ｐｈａｓｅ）において一貫性のチェックを行う場合に主サイトＤＢＳ$_P$ に問い合わせるだけで済み、他の複製データにＷｌｏｃｋをかけにいく必要はなくなる。切断していたサイトが再接続した場合の制御も同じで良い。ＰＴＯ方式において複数種類の複製データにアクセスするトランザクションにあっては、ローカルサイトにそれらの複製を作成してから実行を始めることとする。

【００４９】実施の形態１に係る分散型データベースシステムの一貫性管理方法を超多地点にわたる分散型データベースシステムに適用することを想定した場合には、重み付き投票方法よりＰＴＯ方式の方が優れていると考えられる。なぜなら、ＰＴＯ方式の場合、一貫性のチェックを行う場合に主サイトＤＢＳ$_P$ に問い合わせるだけで済むため、主サイトＤＢＳ$_P$ がダウンしない限り、通信コストの削減を図ることができるからである。

【００５０】そこで、実施の形態１に係る分散型データベースシステムの一貫性管理方法においては、レベルＡの一貫性管理方法としてＰＴＯ方式を採用することにし、以下にレベルＡの一貫性管理方式を具体的に説明する。ここでは、サイトＤＢＳ$_I$ でｒｅａｄ－ｏｎｌｙまたはｒｅａｄ－ｏｎｌｙ以外の更新を含むトランザクションＴが発生するものとする。

【００５１】なお、以下の説明において、複製データはオブジェクトＯに対してｎ個あると仮定し、それをO_1，・・・，O_n と表すと共に、主サイトＤＢＳ$_P$ が有するオブジェクトＯの複製データをO_0 と表す。オブジェクトＯの最後に更新された時刻を表すタイムスタンプをｔ（Ｏ）とし、またトランザクションＴが保持するオブジェクトＯの最終更新タイムスタンプをｔ（Ｏ，Ｔ）と表すことにする。

【００５２】サイトＤＢＳ$_I$ は、トランザクションＴが発生すると、そのトランザクションＴに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションＴがレベルＡの場合、レベルＡの一貫性管理方法が実行される。

【００５３】（１）更新を含むトランザクションの場合

① 読み出し相

サイトＤＢＳ$_I$ は、ローカルデータ、即ち複製データO_i にアクセスし、複製データO_i をプライベート領域（ワークエリア）に読み出す。そして、プライベート領域に読み出された複製データO_i についてｒｅａｄおよびｗｒｉｔｅが行われる。複製データO_i をプライベート領域に読み出す際には、複製データO_i の最後に更新された時刻を表すタイムスタンプｔ（O_i）をトランザクションＴが保持する複製データO_i の最終更新タイムスタンプに設定し、ｔ（O_i，Ｔ）とする。

【００５４】② 確認相

サイトＤＢＳ$_I$ は、主サイトＤＢＳ$_P$ に対してレベルＡの一貫性管理方法の要求を含むＲＥＱＵＥＳＴメッセージを送信する。図４は、ＲＥＱＵＥＳＴメッセージを示す説明図であり、図４に示すデータがＲＥＱＵＥＳＴメッセージに設定され、主サイトＤＢＳ$_P$ に対して送信される。なお、ＲＥＱＵＥＳＴメッセージには、メッセージの送信元であるサイトＤＢＳのサイトＩＤ，トランザクションＩＤ，さらに、トランザクションＴがアクセスした複製データＯｉ毎に、操作（ｒｅａｄまたはｗｒｉｔｅ），対象オブジェクトＩＤ，タイムスタンプｔ（Ｏｉ，Ｔ）および更新後の値が設定される。

【００５５】主サイトＤＢＳ$_P$ は、ＲＥＱＵＥＳＴメッセージを受信し、トランザクションＴにレベルＡが設定されていると判定して、受信したＲＥＱＵＥＳＴメッセージを待ち行列に入力する。待ち行列に入力されたＲＥＱＵＥＳＴメッセージは、入力された順番で処理される。

【００５６】ここで、主サイトＤＢＳ$_P$ は、サイトＤＢＳ$_I$ によってアクセスされた複製データO_i の全てについて、

【数１】

$$t(O_i,T) \geq t(O_0)$$

の条件を満たすか否か危険領域（ｃｒｉｔｉｃａｌ　ｓｅｃｔｉｏｎ）内で判定する。

【００５７】上記条件を満たすと判定した場合、サイトＤＢＳ$_I$ で更新された複製データO_i に該当する複製データO_0 をＲＥＱＵＥＳＴメッセージ中の更新後の値を用いて更新し（更新後の値を保持する）、そのタイムスタンプをｔ（O_0）＝Ｃｌｏｃｋ（現在時刻のタイムスタンプ）に変更した後、図５に示すコミットメッセージを全てのサイトＤＢＳ（サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$。：更新された複製データＯｉに該当する複製データを有するサイトＤＢＳ）に送信する。このコミットメッセージは、ＲＥＱＵＥＳＴメッセージに基づいて生成され、ＲＥＱＵ

ＥＳＴメッセージの送信元であるサイトＤＢＳのサイトＩＤ，トランザクションＩＤ，さらに、複製データＯｉ毎に、対象オブジェクトＩＤ，タイムスタンプｔ（O_0）および更新後の値が設定される。

【００５８】すなわち、更新を含むトランザクションの場合、主サイトＤＢＳ$_r$はそのトランザクション内でサイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$の全てに更新情報の伝播を行う。ただし、２相コミット方式とは異なり、サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$からの返事は必要としない。つまり、ダウンしているサイトＤＢＳがあっても無視して処理を進めることができる。

【００５９】なお、ダウンしているサイトＤＢＳが復旧した場合、そのサイトＤＢＳは、主サイトＤＢＳ$_r$に対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトＤＢＳ$_r$に対して最新の更新状態を問い合わせることができるようにすることもできる。

【００６０】一方、上記条件を満たさないと判定した場合、図６に示すアボートメッセージをサイトＤＢＳ$_1$に送信する。このアボートメッセージは、ＲＥＱＵＥＳＴメッセージに基づいて生成され、ＲＥＱＵＥＳＴメッセージの送信元であるサイトＤＢＳのサイトＩＤ，トランザクションＩＤ，さらに、複製データＯｉ毎に、対象オブジェクトＩＤ，タイムスタンプｔ（O_0）および現在の値O_0が設定される。

【００６１】③　書き込み相

サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$は、コミットメッセージを受信した場合、コミットメッセージに含まれている更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。

【００６２】一方、サイトＤＢＳ$_1$は、アボートメッセージを受信した場合、トランザクションＴをアボートし、アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データO_1およびそのタイムスタンプｔ（O_1）を更新する。

【００６３】（２）ｒｅａｄ－ｏｎｌｙのトランザクションの場合

続いて、ｒｅａｄ－ｏｎｌｙのトランザクションの場合について説明する。サイトＤＢＳ$_1$は、ローカルデータ、即ち複製データO_1にアクセスし、複製データO_1をプライベート領域（ワークエリア）に読み出す。ここで、複製データO_1をプライベート領域に読み出す際には、複製データO_1のタイムスタンプｔ（O_1）をトランザクションＴが保持する複製データO_1のタイムスタンプに設定し、ｔ（O_1，Ｔ）とする。

【００６４】サイトＤＢＳ$_1$は、主サイトＤＢＳ$_r$に対してレベルＡの一貫性管理方法の要求を含むＲＥＱＵＥＳＴメッセージを送信する。このＲＥＱＵＥＳＴメッセージは図４に示したものと同様であるが、ｒｅａｄ－ｏｎｌｙのトランザクションであるため、更新後の値は除かれる。

【００６５】主サイトＤＢＳ$_r$は、ＲＥＱＵＥＳＴメッセージを受信し、トランザクションＴにレベルＡが設定されていると判定して、受信したＲＥＱＵＥＳＴメッセージを待ち行列に入力する。待ち行列に入力されたＲＥＱＵＥＳＴメッセージは順番に処理される。

【００６６】主サイトＤＢＳ$_r$は、サイトＤＢＳ$_1$によってアクセスされた複製データO_1について、

【数２】

$$t(O_i,T) \geq t(O_0)$$

の条件を満たすか否か危険領域（ｃｒｉｔｉｃａｌ　ｓｅｃｔｉｏｎ）内で判定する。

【００６７】上記条件を満たすと判定した場合、図５に示したコミットメッセージをサイトＤＢＳ$_1$に送信する。ただし、ｒｅａｄ－ｏｎｌｙのトランザクションであるため、更新後の値は除かれる。一方、上記条件を満たさないと判定した場合、図６に示したアボートメッセージをサイトＤＢＳ$_1$に送信する。

【００６８】サイトＤＢＳ$_1$は、コミットメッセージを受信した場合、厳格に一貫性が保たれているとしてｒｅａｄ－ｏｎｌｙのトランザクションをコミットする。一方、サイトＤＢＳ$_1$は、アボートメッセージを受信した場合、ｒｅａｄ－ｏｎｌｙのトランザクションをアボートし、アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データO_1およびそのタイムスタンプｔ（O_1）を更新する。

【００６９】（レベルＢの一貫性管理方法）つぎに、レベルＢによる一貫性管理方法について説明する。レベルＡでは一つのトランザクションの中で全てのサイト（サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$）への更新を行うことにしているが、レベルＢでは、別のトランザクションで更新を伝播させることにする。つまり、即時に更新を伝播させるのではなく、後述する各種の差異限定制約条件（本発明の更新伝播条件に該当する）を設け、その条件が満足された場合に更新伝播を行うという処理を行う。

【００７０】ここでは、サイトＤＢＳ$_1$でｒｅａｄ－ｏｎｌｙまたはｒｅａｄ－ｏｎｌｙ以外の更新を含むトランザクションＴが発生するものとして、レベルＢの一貫性管理方法を具体的に説明する。

【００７１】サイトＤＢＳ$_1$は、トランザクションＴが発生すると、そのトランザクションＴに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションＴがレベルＢの場合、レベルＢの一貫性管理方法が実行される。

【００７２】（１）更新を含むトランザクションの場合

① 読み出し相
サイトＤＢＳ$_1$は、ローカルデータ、即ち複製データＯ$_1$にアクセスし、複製データＯ$_i$をプライベート領域（ワークエリア）に読み出す。そして、プライベート領域に読み出された複製データＯ$_i$についてｒｅａｄおよびｗｒｉｔｅが行われる。ここで、複製データＯ$_i$をプライベート領域に読み出す際には、複製データＯ$_i$の最後に更新された時刻を表すタイムスタンプｔ（Ｏ$_i$）をトランザクションＴが保持する複製データＯ$_i$の最終更新タイムスタンプに設定し、ｔ（Ｏ$_i$，Ｔ）とする。

【００７３】② 確認相
サイトＤＢＳ$_1$は、主サイトＤＢＳ$_r$に対してレベルＢの一貫性管理方法の要求を含むＲＥＱＵＥＳＴメッセージを送信する。このＲＥＱＵＥＳＴメッセージは、図４に示したものと同様のものである。

【００７４】主サイトＤＢＳ$_r$は、ＲＥＱＵＥＳＴメッセージを受信し、トランザクションＴにレベルＢが設定されていると判定して、受信したＲＥＱＵＥＳＴメッセージを待ち行列に入力する。待ち行列に入力されたＲＥＱＵＥＳＴメッセージは、入力された順番で処理される。

【００７５】ここで、主サイトＤＢＳ$_r$は、サイトＤＢＳ$_1$によってアクセスされた複製データＯ$_i$の全てについて、

【数３】

$$t(O_i,T) \geq t(O_0)$$

の条件を満たすか否か危険領域（ｃｒｉｔｉｃａｌ ｓｅｃｔｉｏｎ）内で判定する。

【００７６】上記条件を満たすと判定した場合、サイトＤＢＳ$_1$で更新された複製データＯ$_i$に該当する複製データＯ$_0$をＲＥＱＵＥＳＴメッセージ中の更新後の値に更新し（更新後の値を保持する）、そのタイムスタンプをｔ（Ｏ$_0$）＝Ｃｌｏｃｋ（現在時刻のタイムスタンプ）に変更した後、図５に示したコミットメッセージをサイトＤＢＳ$_1$に送信する。加えて、主サイトＤＢＳ$_r$は、上記トランザクションで更新された更新情報をキューに入れておき、差異限定制約条件が満足された場合に、この更新情報を別のトランザクションにおいて他のサイトＤＢＳ（サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$）に送信することによって更新伝播を行う。なお、ＲＥＱＵＥＳＴメッセージを送信して来たサイトＤＢＳ（ここではサイトＤＢＳ$_1$）に対しては、通信コストを下げるために更新伝播を行わないようにする。

【００７７】一方、上記条件を満たさないと判定した場合、図６に示したアボートメッセージをサイトＤＢＳ$_1$に送信する。

【００７８】ここで、上記差異限定制約条件について説明する。差異限定制約条件としては、例えば、以下のａ）～ｅ）等を挙げることができる。

【００７９】ａ）リフレッシュする周期を設定し、設定した周期毎に更新伝播を行う。主サイトＤＢＳ$_r$は、一定の周期毎に、他の全てのサイトＤＢＳ（サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$）に更新情報を送信することにより、更新伝播を行う。主サイトＤＢＳ$_r$は、データ更新情報をキューに入れておき、周期毎にそれを送り出す。受け側のサイトＤＢＳ（サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$）はこの更新情報が送られてくることにより、主サイトＤＢＳ$_r$がダウンしていないことを知ることができる。

【００８０】ｂ）キューにたまっている更新情報の数が一定数を超えた場合に更新伝播を行う。例えば限度数を３とした場合は、更新情報がキューに３つたまったところで更新伝播が行われる。

【００８１】ｃ）更新伝播を行う基準となるデータの領域を決めておき、その領域のデータが更新された場合に更新伝播を行う。例えば、書誌情報のうち、タイトルおよび作者の２つの属性が更新された場合は即時に更新伝播を行うというものであり、直ちに更新情報を公知にする必要があるときに有効である。スキーマの属性として、オブジェクトやその属性に即時更新伝播フラグをつけておき、それがＯＮになっている情報に更新があった場合は即時に更新伝播を行うようにする。ただし、この条件を設定する場合には、他の条件と共に設定する必要がある。

【００８２】ｄ）更新伝播を行う基準となる閾値を設定し、設定した閾値を超えた場合に更新伝播を行う。データの更新の度合いに対して閾値を設定しておく。例えば、コンテンツの１／２０以上が更新された場合に更新伝播するというような絶対的な値を設定し、または、参照された回数が１００を超えたら即時更新伝播するというような設定を行う。後者の場合、本や番組で人気が出たものはそれを参照する人が多いことを示しているため、即時に情報行進を行うほうが良いという考えに基づくものである。

【００８３】ｅ）バージョンが大きく変わった場合に更新伝播を行う。本の改訂版等でマイナーな誤字の直し程度であれば更新伝播を待つようにするが、大幅な修正が加えられた改訂版が出たような場合には、直ちに更新伝播を行うことにするというものである。

【００８４】なお、上記ａ）～ｅ）の差異限定制約条件は組み合わせて設定することができる。

【００８５】次に、差異限定制約条件を用いた更新伝播手順について説明する。図７（ａ）は差異限定制約条件を設定する手順を示すフローチャートである。図７（ａ）に示すように、システムのセットアップ時において、上記ａ）～ｅ）に示した差異限定制約条件を単独または組み合わせて主サイトＤＢＳ$_r$に設定する（Ｓ７００：本発明の更新伝播設定工程に該当する）。

【００８６】なお、差異限定制約条件を設定する際には、主サイトＤＢＳｒを操作して設定しても良いし、サイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎから設定することもできる。

【００８７】また、差異限定制約条件は、上記ステップＳ７００の処理を再び行うことによって後に変更することも可能であるが、図３（ａ）に示したステップＳ３００において一貫性管理のレベルを設定する際にいずれかの差異限定制約条件を設定しておく必要がある。

【００８８】さらに、サイトＤＢＳ毎に異なる差異限定制約条件を設定することもできる。例えば、サイトＤＢＳ１については差異限定制約条件ａ）で更新伝播を行い、サイトＤＢＳ２については差異限定制約条件ｂ）で更新伝播を行うように主サイトＤＢＳｒに対して設定することもできる。この場合、主サイトＤＢＳｒは、各サイトＤＢＳ毎に設定された差異限定制約条件を管理し、各サイト毎に異なる差異限定制約条件を用いて更新伝播を行うことになる。

【００８９】図７（ｂ）は図７（ａ）に示すようにして設定した差異限定制約条件に基づいて更新伝播を行う手順を示すフローチャートである。主サイトＤＢＳｒは、常にまたは定期的にキューに入れられた更新情報について、差異限定制約が満足されたか否かを判定する（Ｓ７０１）。

【００９０】そして、ステップＳ７０１において差異限定制約が満足されたと判定した場合には、キューに入れられた更新情報をサイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎに送信することによって更新伝播処理を行う（Ｓ７０２）。

【００９１】③　書き込み相

サイトＤＢＳｉは、コミットメッセージを受信した場合、コミットメッセージに含まれている更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。

【００９２】また、サイトＤＢＳ１，サイトＤＢＳ３，・・・，サイトＤＢＳｎは、主サイトＤＢＳｒから更新伝播を受信した場合、更新後の値およびそのタイムスタンプ値を用いて該当する複製データを更新する。この際、２相コミット方式とは異なり、サイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎからの返事は必要としない。つまり、ダウンしているサイトがあっても無視して処理を進めることができる。

【００９３】なお、ダウンしているサイトＤＢＳが復旧した場合、そのサイトＤＢＳは、主サイトＤＢＳｒに対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトＤＢＳｒに対して最新の更新状態を問い合わせることができるようにすることもできる。

【００９４】一方、サイトＤＢＳｉは、アボートメッセージを受信した場合、トランザクションをアボートし、アボートメッセージに含まれている最新値およびタイムスタンプ値を用いて複製データＯｉおよびそのタイムスタンプｔ（Ｏｉ）を更新する。

【００９５】（２）ｒｅａｄ－ｏｎｌｙのトランザクションの場合

レベルＢの場合は、差異限定制約条件に基づいて更新伝播を行うため、主サイトＤＢＳｒを除き、サイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎの複製データは最新バージョンではない可能性が常にある。それをｒｅａｄ－ｏｎｌｙのトランザクションにおいて許可するかしないかでレベルＡとレベルＢとの差が生じる。

【００９６】レベルＢにおけるｒｅａｄ－ｏｎｌｙのトランザクション場合では、ローカルサイト（サイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎ）の複製データをそのままｒｅａｄするようにして、直列可能性チェックを緩和する。つまりｒｅａｄ－ｏｎｌｙのトランザクションであれば、他のサイトＤＢＳとの調整を行うことなくローカルに実行してしまうことにする。このようにした場合であっても、１コピー直列化は保証される。つまり、他の更新が起こる前にｒｅａｄしていたと考えることで直列化できる。

【００９７】（レベルＣの一貫性管理方法）レベルＢによる一貫性管理方法は、更新を含むトランザクションに対しては厳格、つまり強い一貫性管理を採用している。これに対し、レベルＣによる一貫性管理方法においては、システム全体の処理効率を更に向上させるため、更新を含むトランザクションも緩和して弱い一貫性管理を行うことにする。

【００９８】そこで、レベルＣの一貫性管理方法を説明する。ここでは、サイトＤＢＳｉでｒｅａｄ－ｏｎｌｙまたはｒｅａｄ－ｏｎｌｙ以外の更新を含むトランザクションＴが発生するものとする。

【００９９】サイトＤＢＳｉは、トランザクションＴが発生すると、そのトランザクションＴに設定されている一貫性管理のレベルを判定する処理を行う。発生したトランザクションＴがレベルＣの場合、レベルＣの一貫性管理方法が実行される。

【０１００】（１）更新を含むトランザクションの場合

①　読み出し相

サイトＤＢＳｉは、ローカルデータ、即ち複製データＯｉにアクセスし、複製データＯｉをプライベート領域（ワークエリア）に読み出す。そして、プライベート領域に読み出された複製データＯｉについてｒｅａｄおよびｗｒｉｔｅが行われる。ここで、レベルＣの一貫性管理方法においては、弱い一貫性管理を採用することにより、主サイトＤＢＳｒを介することなく、サイトＤＢＳ１，サイトＤＢＳ２，サイトＤＢＳ３，・・・，サイトＤＢＳｎ毎に単独で複製データの更新を行うことができる。そのため、サイトＤＢＳｉは、読み出した複製デー

タO_iを更新し、更新を含むトランザクションをコミットする。

【０１０１】② 確認相

サイトＤＢＳ$_i$は、上記コミット時に主サイトＤＢＳ$_r$宛てに承認リクエストメッセージ（更新ログ）を送信する。承認リクエストメッセージは、図４に示したＲＥＱＵＥＳＴメッセージと同様のものである。

【０１０２】主サイトＤＢＳ$_r$は、サイトＤＢＳ$_i$からの承認リクエストメッセージに対して承認または非承認を与える。承認・非承認の判定条件はＰＴＯ方式の中で説明した通常のタイムスタンプ方式によるものである。

【０１０３】ここで、承認・非承認を判定する際に、サイトＤＢＳ$_i$の承認リクエストメッセージと競合する他のサイトＤＢＳの承認リクエストメッセージが存在しない場合には、サイトＤＢＳ$_i$において行われた更新が承認される。その結果、主サイトＤＢＳ$_r$の該当する複製データが更新される（更新後の値を保持し、タイムスタンプを変更）と共に、レベルＢで説明したように、その更新情報がキューに入れられ、差異限定制約条件が満足された場合に、この更新情報をサイトＤＢＳ（サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$。）に送信することによって更新伝播が行われる。なお、差異限定制約条件については、レベルＢで説明した通りであるため、ここでは説明を省略する。また、レベルＢの場合と同様に、承認リクエストメッセージを送信して来たサイトＤＢＳ（ここではサイトＤＢＳ$_i$）に対しては、通信コストを下げるために更新伝播を行わないようにする。

【０１０４】一方、例えば、サイトＤＢＳ$_i$およびサイトＤＢＳ$_j$の承認リクエストメッセージが競合したものとする。具体的には、サイトＤＢＳ$_i$およびサイトＤＢＳ$_j$で同一の複製データO_iを保持していて、両方のトランザクションで同時刻に複製データO_iを更新し、両方のトランザクションがコミットされた後、承認リクエストメッセージ（更新ログ）が主サイトに届いたものとする。主サイトＤＢＳ$_r$は、承認リクエストメッセージが先に届いたサイトの更新を承認し、後から届いた方の更新は承認しない。ここで、サイトＤＢＳ$_j$の承認リクエストメッセージがサイトＤＢＳ$_i$の承認リクエストメッセージより先に主サイトＤＢＳ$_r$に届いたものとすると、サイトＤＢＳ$_i$の更新は承認されないことになる。この場合、主サイトＤＢＳ$_r$は、サイトＤＢＳ$_i$に対してアボートメッセージを送信する。

【０１０５】③ 書き込み相

更新が承認された場合、サイトＤＢＳ$_i$は、既に複製データを更新するためのトランザクションはコミットされているため何ら処理を行う必要はない。一方、他のサイトＤＢＳは、差異限定制約条件によって更新情報が伝播されて来るため、その更新情報に基づいて複製データを更新する。この際、２相コミット方式とは異なり、サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$からの返事は必要としない。つまり、ダウンしているサイトがあっても無視して処理を進めることができる。

【０１０６】なお、ダウンしているサイトＤＢＳが復旧した場合、そのサイトＤＢＳは、主サイトＤＢＳ$_r$に対して何らかのトランザクションを実行することにより、最新の更新状態を知ることができる。また、復旧した後に、主サイトＤＢＳ$_r$に対して最新の更新状態を問い合わせることができるようにすることもできる。

【０１０７】一方、アボートメッセージを受信した場合は、既にトランザクションはコミットにより完了してしまっているため通常のアボートを行うことはできない。そこで、レベルＣにおいては、他のサイトＤＢＳから必要なバージョンの複製データをコピーすることにより更新してしまった複製データを回復することにする。

【０１０８】ここで、上記承認リクエストメッセージ中の複製データのうち、主サイトＤＢＳ$_r$から更新されている値をコピーする処理、つまりトランザクション自身がアボート用にデータを格納しておき、このデータを用いて更新してしまった複製データを回復するという処理も考えられる。しかしながら、マルチメディアデータは一般にサイズが大きいため、トランザクションの中で複製を作ることはコストがかかり問題となる。また、更新頻度の小さい応用においては更新が衝突する可能性が少ないため、そのような応用においてアボート用にデータの複製を作ることは効率的ではない。

【０１０９】よって、アボートメッセージを受信した場合の処理として、記憶容量およびコピーに要する時間の点からも他のサイトＤＢＳから必要なバージョンの複製データをコピーして回復する処理を行う方が効率的であるといえる。

【０１１０】（２）ｒｅａｄ－ｏｎｌｙのトランザクションの場合

レベルＣにおけるｒｅａｄ－ｏｎｌｙのトランザクション場合では、ローカルサイト（サイトＤＢＳ$_1$，サイトＤＢＳ$_2$，サイトＤＢＳ$_3$，・・・，サイトＤＢＳ$_n$。）の複製データをｒｅａｄするようにして、直列可能性チェックを緩和する。つまりｒｅａｄ－ｏｎｌｙのトランザクションであれば、他のサイトとの調整を行うことなくローカルに実行してしまうことにする。このようにした場合であっても１コピー直列化は保証される。つまり、他の更新が起こる前にｒｅａｄしていたと考えることで直列化できる。

【０１１１】ここで、レベルＣの応用例として、検索エージェント（人間に代わって複数のＤＢサイトを縦断検索してくれるようなプログラム）のＤＢを考える。従来のデータベースはオブジェクト値を永続的に格納することを目的としていた。ところが、検索エージェントのような応用を考えた場合、その記憶としてデータベースを

使うならば、永続的というよりは「少々間違っていても良く、間違っていることが判定できれば良い。そして、間違っていたら正しい値に更新すれば良い。」という特徴を持つ。検索サイト名等を格納している場合において、もし、そのサイトが移動したとしても、そのサイトにアクセスしたときに初めて「そのサイトが移動していたこと」がわかれば良い。人間の記憶の変わりにＤＢを利用とする場合、このような「多少間違ったデータでも良い」という応用例を多く見出すことができる。

【０１１２】なお、以上説明した一貫性管理のレベルＡ～Ｃの特徴を一つにまとめると、図８のようになる。

【０１１３】このように、実施の形態１に係る分散型データベースシステムの一貫性管理方法によれば、分散型データベースシステムの一貫性管理のレベルとしてレベルＡ～Ｃを設け、トランザクションの種類に応じてレベルＡ～Ｃによる異なる一貫性管理を行うことにより、複製データの一貫性管理における利便性の向上を図ることができる。

【０１１４】また、実施の形態１に係る分散型データベースシステムの一貫性管理方法を用いることにより、それぞれ異なる一貫性管理のレベルでトランザクションを実行することができ、トランザクションの種類に応じて最適な複製データの一貫性管理を行うことができる。

【０１１５】さらに、レベルＢおよびレベルＣの場合は、差異限定制約が満足された場合に更新伝播を行うことにしたことにより、更新がある毎に更新伝播を行う必要をなくすことができ、複製データの一貫性保持のコストを削減することができる。

【０１１６】〔実施の形態２〕つぎに、実施の形態２に係る分散型データベースシステムの一貫性管理方法について説明する。例えば、電子図書館等の応用においては、新規の書誌データを一括して登録することが多く、また、大学等でも新入生や新人のデータ登録を一括して行うことが多い。こうした新規登録トランザクションは、既にあるデータを参照しないで独立に処理することが可能であることから、他のトランザクションとの競合の確認処理を省略しても何ら問題は発生せず、むしろ処理の効率化を図る上で好ましい。そのため、実施の形態２に係る分散型データベースシステムの一貫性管理方法においては、実施の形態１で説明した処理に加えて、一貫性管理のレベルを判定する際に新規登録トランザクションか否かを判定し、新規登録トランザクションである場合には、他のトランザクションとの競合の確認処理を省略して、そのまま登録処理を行うことができるようにするものである。

【０１１７】ところで、楽観的制御においては、読み出し相，確認相および書き込み相からなる処理を行う必要がある。ところが新規登録のみのトランザクションでは、他のトランザクションと競合する可能性は全くないため、トランザクションが競合するか否かの確認を省略し、読み出し相の処理が終わった後、直接書き込み相の処理を実行することができる。

【０１１８】そこで、実施の形態２に係る分散型データベースシステムの一貫性管理方法においては、既存のデータと独立なデータを新規登録する場合について、独自のトランザクションレベル（例えば、新規登録処理フラグを設定する）を設け、そのレベルのトランザクションについては一切競合チェックを行わないことにする。

【０１１９】新規登録トランザクションは、実施の形態１で説明したレベルＡ～Ｃのいずれ対しても適用することができる。ただし、効率を考えた場合には、各サイトＤＢＳ単位で更新を行うことができるレベルＣに適用することが最も優れている。そのため、実施の形態２においては、この新規登録トランザクションをレベルＣに適用することを前提として、処理の概略を説明する。図９（ａ）は一貫性管理のレベルおよび新規登録トランザクションを設定する手順を、図９（ｂ）は図９（ａ）に示すようにして設定した一貫性管理のレベルおよび新規登録トランザクションの設定に基づいて一貫性管理方法および新規登録を実行する手順を示すフローチャートである。

【０１２０】まず、図９（ａ）に示すように、トランザクションの種類に応じて一貫性管理のレベルを設定し（実施の形態１参照）、かつ、データの新規登録のためのトランザクションを新規登録トランザクションとして設定する（Ｓ９００）。

【０１２１】図９（ａ）のステップＳ９００で新規登録トランザクションについての一貫性管理レベルを設定した後、図９（ｂ）に示すように、分散型データベースシステム内にトランザクションが発生すると（Ｓ９０１）、発生したトランザクションに設定されている一貫性管理のレベルの判定処理およびトランザクションが新規登録トランザクションか否かの判定処理を実行する（Ｓ９０２：本発明の判定工程に該当する）。

【０１２２】そして、ステップＳ９０２において、トランザクションが新規登録トランザクションであると判定した場合、設定されているレベルに関係なく、レベルＣと判定する（Ｓ９０３：本発明の判定工程に該当する）。そして、ステップＳ９０６に進み、レベルＣの一貫性管理方法を実行する。

【０１２３】なお、レベルＣで新規登録トランザクションを実行する場合においては、他のサイトＤＢＳにおけるトランザクションとの競合の可能性がないことから、承認リクエストメッセージを用いた競合のチェックは省略される。

【０１２４】このように、実施の形態２に係る分散型データベースシステムの一貫性管理方法によれば、データの新規登録を行うためのトランザクションを実行する場合において、他のトランザクションとの競合をチェックする処理を省略することにしたため、データを新規登録

する際の処理効率を向上させることができる。

【０１２５】ただし、ＩＮＳＥＲＴ操作の中には既存のデータと意味的に関連付けられているものがある。例えば、ｒｅａｄした２つの値の平均値を求め、その値を使ってＩＮＳＥＲＴするようなトランザクションの場合である。このような場合には、既存のデータと意味的に関連があるため、そのトランザクションは新規登録トランザクションではない。

【０１２６】〔実施の形態３〕上述したレベルＢおよびレベルＣの一貫性管理方法においては、主サイトＤＢＳ，に予め設定された差異限定制約条件が満足された場合に、主サイトＤＢＳ，から各サイトＤＢＳに対して更新情報を送信し、更新伝播を行うことにしている（図７参照）。実施の形態３においては、更新伝播を行うタイミングとして、主サイトＤＢＳ，に設定された差異限定制約条件が満足された場合に加え、トランザクションが発生したサイトＤＢＳ側でコンテンツの内容に応じて更新伝播を行うタイミングを設定することができるようにする。このようにするのは、コンテンツによっては早期に公開すべきものや多少公開が遅れても良いとされるものがあり、例えば、台風情報等のように早期に公開すべき情報については５分以内に全てのサイトＤＢＳに更新伝播を完了したいという要請に答えることができるようにするためである。

【０１２７】図１０は、本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。図１０に示す分散型データベースシステムは、図２に示した構成に加えて、更新伝播を行う際に、衛星１０００を用いた放送を利用して主サイトＤＢＳ，から各サイトＤＢＳに更新情報を配信できるようにしたものである。一般に衛星１０００等を利用して情報を配信することは例えばインターネットを利用する場合と比べてコストが高いが、緊急に全てのサイトＤＢＳに更新伝播を行いたい場合等において放送により更新情報を配信することにより、全てのサイトＤＢＳに対して更新情報を短時間にかつ一度に配信することができるという利点がある。

【０１２８】なお、図１０において、１００１ａ～１００１ｅは、放送波を用いて更新情報を送信または受信するためのアンテナを、１００２は実施の形態１および２で既に説明した各種情報を送受信するための通信回線をそれぞれ示している。

【０１２９】また、図１０には、衛星１０００を用いた放送により情報やり取りを行うことができる分散型データベースシステムを一例として示したが、これに加えて、または代えて地上波等、情報を送信可能な他の手段を用いることにしても良い。

【０１３０】次に、実施の形態３に係る分散型データベースシステムの一貫性管理方法について、サイトＤＢＳ，でレベルＣの一貫性管理方法が設定されたトランザクションが発生した場合を例にとって説明する。なお、ここではレベルＣの場合について説明するが、レベルＢの場合にも同様に適用できることは明らかである。また、実施の形態１および２で既に説明した事項については、簡単に説明することにする。

【０１３１】図１１は、更新伝播タイミング設定処理を示すフローチャートである。サイトＤＢＳ１のユーザは、複製データの更新を行う際に、例えば、予め用意された入力欄等に以下に説明する情報を設定する（Ｓ１１０１）。

【０１３２】ステップＳ１１０１においては、緊急度，公開指定日時，伝播遅延許容日時，伝播方式が設定されるものとする。緊急度には例えば以下のようなレベルがあり、これらのレベルは予めシステム構築者が設定する。したがって、ユーザは設定されている緊急度レベルの中から所望のレベルを選択し、緊急度の項目に設定するすることになる。

【０１３３】
・緊急度１：１０分以内に更新伝播を行う。
・緊急度２：６０分以内に更新伝播を行う。
・緊急度３：１日以内に更新伝播を行う。
・緊急度４：３日以内に更新伝播を行う。

【０１３４】公開指定日時の項目は、例えば報道ニュース等には明朝１０時までは情報を公開してはならないという要求があることから設けられたものである。したがって、ユーザはこの項目に対して所望の日時を設定する。なお、この公開指定日時が指定された場合、主サイトＤＢＳ，は、公開指定日時が経過した後に更新された複製データに関する情報について更新伝播を行うようにする。また、これに代えて、公開指定日時の経過前に更新伝播を行うことにしても良い。いずれを用いるかについては、後述する更新伝播のスケジュールを設定する際（図１３のステップＳ１３０４参照）に決定すれば良い。ただし、公開指定日時の経過前に更新伝播を行うことにする場合、各サイトＤＢＳにおいて公開指定日時が経過する前に該当する更新後のデータを公開しないようにする手法を確立しておくことが必要となる。

【０１３５】伝播遅延許容日時の項目には、更新伝播の遅れが許される最大限の日時を設定する。

【０１３６】さらに、配信方式の項目には、通信（通信回線１００２を用いる），衛星放送，地上波放送等のいずれの方式を用いて更新伝播を行うかを設定する。

【０１３７】そして、ここではレベルＣであるため、サイトＤＢＳ１においてトランザクションをコミットするかアボートするかが判定され、コミットの場合はサイトＤＢＳ，の該当する複製データが更新される。なお、実施の形態２で説明した新規登録トランザクションの場合には、他のトランザクションとの競合のチェック処理は省略される。

【０１３８】コミットの場合、サイトＤＢＳ$_l$は、上述したように承認リクエストメッセージを主サイトＤＢＳ$_r$に通信回線１００２を介して送信するが、その際、図１２（ａ）および図１２（ｂ）に例として示すデータを承認リクエストメッセージ（図４に示したＲＥＱＵＥＳＴメッセージと同一内容：レベルＢの場合も同様）に添付する。その結果、図１１のステップＳ１１０１で設定した緊急度等は、属性値として承認リクエストメッセージに設定されることになる。

【０１３９】主サイトＤＢＳ$_r$は、サイトＤＢＳ$_l$から承認リクエストメッセージを受信すると、他のサイトＤＢＳで行われた更新処理と競合しないか否かを判定し、競合しないと判定した場合は以下の更新伝播管理処理を実行する。なお、新規登録トランザクションの場合は競合のチェック処理は省略される。また、競合のチェック処理と並行して更新伝播管理処理を実行することにしても良いが、他のサイトの更新処理と競合すると判定された場合は更新伝播管理処理を途中で終了することになる。

【０１４０】図１３は、主サイトＤＢＳ$_r$による更新伝播管理処理を示すフローチャートである。主サイトＤＢＳ$_r$は、サイトＤＢＳ$_l$から受信した承認リクエストメッセージから依頼元ＩＤ（サイトＩＤ），依頼日時，更新ＯＩＤリスト，緊急度，公開指定日時，更新遅延許容日時および配信方式をコピーし（Ｓ１３０１）、さらに、依頼到着日時（主サイトＤＢＳ$_r$が承認リクエストメッセージを受信した日時），データサイズ（送るべきデータのサイズ）を計測・設定し（Ｓ１３０２）、図１４に示すような更新伝播管理情報を生成する。この更新伝播管理情報は、課金のためのデータとして利用することができる。

【０１４１】なお、ステップＳ１３０２において、承認リクエストメッセージに伝播遅延許容日時が設定されていない場合には、

依頼到着日時＋指定された緊急度の許容時間（緊急度１なら１０分）

という計算を行い、伝播遅延許容日時を設定する。

【０１４２】更新伝播管理情報を生成した後、主サイトＤＢＳ$_r$は、指定された緊急度が許容可能か否かについて、以下の基準に基づいて判定する（Ｓ１３０３）。

【０１４３】

・［依頼到着日時＋指定された緊急度の許容時間］＞伝播遅延許容日時の場合

・［公開指定日時＋指定された緊急度の許容時間］＞伝播遅延許容日時の場合

・（データサイズ×送付サイト数）の値が大きすぎる場合

なお、ここで、送付サイト数は、例えばサイトＤＢＳ$_l$が更新を行った複製データ（オブジェクト）と同一のオブジェクトに対する複製データを有する他のサイトＤＢＳの数であり、どのサイトＤＢＳが送付サイトに該当するかは主サイトＤＢＳ$_r$が判断する。

【０１４４】更新伝播管理情報中の値が上記３つの条件の少なくとも一つを満たすような場合、主サイトＤＢＳ$_r$は、指定された緊急度が許容可能ではないと判定し、許容不可能メッセージを生成してサイトＤＢＳ$_l$（依頼元のサイト）に送信し、図１２に示したデータのみを再度送信するように要求する（Ｓ１３０７）。

【０１４５】そして、サイトＤＢＳ$_l$（依頼元のサイト）からメッセージに対する応答として図１２に示したデータが送信され、そのデータを受信した場合（Ｓ１３０８）、主サイトＤＢＳ$_r$はステップＳ１３０１に戻って再度更新伝播管理処理を実行する。

【０１４６】一方、ステップＳ１３０３において、指定された緊急度が許容可能と判定した場合、主サイトＤＢＳ$_r$は、更新伝播管理情報中の更新遅延許容日時より前に更新情報を各サイトＤＢＳに配信できるように、更新伝播を行うスケジュールを設定する（Ｓ１３０４）。例えば、主サイトＤＢＳ$_r$は、実施の形態１で説明した差異限定制約条件を用いて更新伝搬を行うスケジュールを設定する。いずれの差異限定制約条件を用いるかについては、例えば、図１４に示した更新伝播管理情報中の伝播遅延許容日時に更新伝播が間に合うことを第１の基準とし、これに加えて更新伝播のコストが低いものはどれかを第２の基準として判断する。そして、主サイトＤＢＳ$_r$は、選択した差異限定制約条件を用いて更新伝播を行うスケジューリングを行い、伝播遅延許容日時を守れるか否かを判定し、守れないと判定した場合には、より早い時期に更新伝播を行うことができる差異限定制約条件を用いて再度更新伝播を行うスケジューリングを行う。換言すれば、更新遅延伝播ルールとして各種の差異限定制約条件があるが、指定された緊急度を守ることができない可能性がある差異限定制約条件は採用しないことにする。

【０１４７】その後、主サイトＤＢＳ$_r$は、ステップＳ１３０４で設定したスケジュールに従い、更新伝播を実行する（Ｓ１３０５）。更新伝播は、更新伝播管理情報中に設定されている配信方式に従って更新情報を各サイトＤＢＳに配信する。例えば、配信方式として「衛星」が設定されている場合は、衛星１０００を用いて放送により更新情報を各サイトＤＢＳに配信する。また、配信方式として「通信」が設定されている場合は、通信回線１００２を介して更新情報を各サイトＤＢＳに配信する。なお、通信回線１００２を介して更新情報を各サイトＤＢＳに配信する場合、承認リクエストメッセージを送信して来たサイトＤＢＳ（ここではサイトＤＢＳ$_l$）に対しては、通信コストを下げるために更新情報の配信は行わないようにする。

【０１４８】そして、主サイトＤＢＳ$_r$は、更新伝播の開始日時および完了日時のログと配信の遅れに関する情

報とを更新伝播管理情報中に設定する（Ｓ１３０６）。配信の遅れに関する情報とは、更新伝播管理情報中の伝播遅延許容日時からどれくらい遅れて配信されたかを示し、

伝播完了日時－伝播遅延許容日時

で計算される。計算した結果が正の場合は、伝播遅延許容日時から遅れて更新伝播が行われたことを示し、その値を更新伝播管理情報中の「配信の遅れ」に設定する。一方、計算した結果が負の場合は、伝播遅延許容日時の前に更新伝播が完了したことを示し、負の場合は配信の遅れは生じていないため、「０」を更新伝播管理情報中の「配信の遅れ」に設定する。

【０１４９】このように、実施の形態３に係る分散型データベースシステムの一貫性管理方法によれば、コンテンツの内容に適したタイミングで更新結果を各サイトＤＢＳに配信することができるため、分散型データベースシステムの利便性の向上を図ることができる。

【０１５０】また、更新結果を各サイトＤＢＳに配信する方法として、衛星１０００を用いた放送による配信方法や通信回線１００２を用いた通信による配信方法等を選択することができるため、緊急度に応じて適切な配信方法を選択でき、分散型データベースシステムの利便性の向上を図ることができる。

【０１５１】なお、詳細な説明については省略するが、レベルＡの場合においても、放送により図４に示したコミットを配信することにしても良い。

【０１５２】以上説明した実施の形態１～３に係る分散型データベースシステムにおいては、トランザクションの種類毎に一貫性管理レベルについてレベルＡ～Ｃのいずれかを設定し、設定したレベルで一貫性管理を行うことにしたが、例えば、レベルＣの一貫性管理レベルのみが利用可能なシステムやレベルＡおよびレベルＣの一貫性管理レベルが利用可能なシステム等、分散型データベースシステムをどのように運用するかによって、様々な設計・変更が可能なことは明らかである。

【０１５３】また、実施の形態１～３に係る分散型データベースシステムの一貫性管理方法は、図３～図５ならびに図１１および図１３に示したフローチャートの手順に従って、予め用意されたプログラムをコンピュータで実行することによって実現される。このプログラムは、ハードディスク，フロッピーディスク，ＣＤ－ＲＯＭ，ＭＯ，ＤＶＤ等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行可能である。また、このプログラムは、上記記録媒体を介して、またはネットワークを介して配布することができる。

【０１５４】〔適用可能な応用分野〕以上説明した実施の形態１～３に係る分散型データベースシステムの一貫性管理方法は、以下に説明するような応用分野に適用することができる。ただし、本発明の分散型データベースシステムの一貫性管理方法を適用する応用分野を以下に説明するものに限定するものではない。

【０１５５】本発明の分散型データベースシステムの一貫性管理方法は、（１）マルチメディアのコンテンツを提供し、その視聴に対して課金を行うシステムであり、（２）コンテンツの更新には時間的遅れを許し、（３）システム内に構築されるＤＢ（データベース）としては、コンテンツＤＢ・インデックス情報ＤＢ（２次情報）・利用者統計情報ＤＢ・課金情報ＤＢという種類を有する分散型データベースシステムに適用することができる。

【０１５６】上記ＤＢの特徴としては、以下のように考えられる。まず、コンテンツＤＢとしては、（１）データのサイズが超大で、データはマルチメディアのＢＬＯＢ（Ｂｉｎａｒｙ　Ｌａｒｇｅ　Ｏｂｊｅｃｔ）であり、（２）複製データが多数必要であり（コンテンツがマルチメディアであるため、サイズが大きく、アクセスコストを下げるためには複製データを多数生成することが必要）、（３）読み出す情報は多少古くても良く、（４）読み出し頻度は大であるものである。（５）更新伝播は遅延しても良い、というものが考えられる。

【０１５７】インデックス情報ＤＢとしては、（１）複製データは多数必要であり、（２）複製データデータのサイズは小～中であり、（３）読み出す情報は多少古くても良く、（４）読み出し頻度は大で、（５）更新伝播は遅延しても良い、というものが考えられる。

【０１５８】利用者統計情報ＤＢおよび課金情報ＤＢとしては、（１）複製データの数は少なくても良く、（２）データのサイズはシステムに依存し、（３）読み出す情報は多少古くても良く、（４）読み出し頻度は小であり、（５）更新伝播は遅延しても良く、集計する側への伝播は遅延しても良い、というものが考えられる。ただし、利用者情報は紛失してもあまり問題とならないが、更新中に課金情報が紛失または欠落することは許されない。

【０１５９】上記分散型データベースシステムにおいて発生する典型的なトランザクションとしては、（１）利用者によるコンテンツ検索（アクセス頻度大）、（２）コンテンツ提供者によるコンテンツの新規追加および修正（更新頻度は小）、（３）コンテンツ提供者によるインデックス情報の新規追加および修正（更新頻度は小）、（４）課金情報の収集（各利用者に対する課金収集は頻度が小さいが、利用者数が膨大であるため、大きくなる）、および、（５）利用統計情報の収集（各利用者に対する課金収集は頻度が小さいが、利用者数が膨大であるため、大きくなる）、が考えられる。

【０１６０】上記条件を満たすシステム例としては、（１）電子図書館システム（コンテンツは書籍，雑誌，ＣＤ－ＲＯＭ，ビデオ等。インデックス情報は書誌情報。）、（２）ビデオ・オン・デマンド（ＶＯＤ）シス

テム（コンテンツはビデオ。インデックス情報はビデオの各種属性。）、（３）カラオケシステム（コンテンツはＭＩＤＩ等の音楽情報，動画情報および歌詞の文字情報等。インデックス情報は曲に関する各種属性。）、等が考えられる。

【０１６１】

【発明の効果】以上説明したように、本発明の分散型データベースシステムの一貫性管理方法（請求項１）によれば、予め複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、一貫性管理のレベルを判定する判定工程と、判定工程で第１レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、判定工程で第２レベルであると判定された場合に、厳格な一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第２の一貫性管理工程と、判定工程で第３レベルであると判定された場合に、弱い一貫性管理方法を用いて一貫性管理を行い、かつ、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、を含むことにより、トランザクションの種類に応じて最適なレベルの一貫性管理を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。

【０１６２】また、本発明の分散型データベースシステムの一貫性管理方法（請求項２）によれば、予め複製データに対するリードまたは更新を含む各トランザクションの種類に、一貫性管理のレベルを表す第１レベル，第２レベルおよび第３レベルのいずれかを設定する一貫性レベル設定工程と、システム内においてトランザクションが発生した場合に、発生したトランザクションの種類に基づいて、一貫性管理のレベルを判定する判定工程と、判定工程で第１レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を即時行う第１の一貫性管理工程と、判定工程で第２レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションは緩和された一貫性管理方法を用いて一貫性管理を行い、更新を含むトランザクションは厳格な一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第２の一貫性管理工程と、判定工程で第３レベルであると判定された場合に、ｒｅａｄ－ｏｎｌｙのトランザクションまたは更新を含むトランザクションのいずれであっても、弱い一貫性管理方法を用いて一貫性管理を行い、第２のサイトから複数の第１のサイトへの更新伝播を遅延させて行う第３の一貫性管理工程と、を含むことにより、トランザクションの種類に応じて最適なレベルの一貫性管理を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。

【０１６３】また、本発明の分散型データベースシステムの一貫性管理方法（請求項３）によれば、請求項１または２に記載の分散型データベースシステムの一貫性管理方法において、第２および第３の一貫性管理工程における更新伝播の遅延について、第２のサイトにおいて予め設定されている更新伝播条件が満足された場合に更新伝播を行うこととしたことにより、更新が起こる毎に更新伝播を行う必要をなくすことができるため、データの一貫性保持のコストを削減することができる。

【０１６４】また、本発明の分散型データベースシステムの一貫性管理方法（請求項４）によれば、請求項１～３のいずれか一つに記載の分散がデータベースシステムの一貫性管理方法において、第２および／または第３の一貫性管理工程における更新伝播の遅延とは、第１のサイトにおいてトランザクション毎に更新伝播を行うタイミングを定めた更新伝播条件が設定され、第２のサイトにおいて更新伝播条件を満足するように複数の第１のサイトへの更新伝播を行うこととしたため、ユーザの所望の更新伝播条件を設定することができる。したがって、ユーザの所望の更新伝播条件を設定することができるため、更新伝播がいつ行われるかを容易に予測することができ、分散型データベースシステムの利便性のさらなる向上を図ることができる。

【０１６５】また、本発明の分散型データベースシステムの一貫性管理方法（請求項５）によれば、請求項４に記載の分散型データベースシステムの一貫性管理方法において、第２および／または第３の一貫性管理工程が、少なくとも通信回線および放送波を用いて第２のサイトから複数の第１のサイトへの更新伝播を行うことができ、更新伝播条件が、通信回線および放送波のいずれを用いて更新伝播を行うかについての指定を含むことにより、更新伝播条件に応じて適切な方法で更新伝播を行うことができるため、分散型データベースシステムの利便性の向上を図ることができる。すなわち、緊急に更新伝播を行う必要がある場合にはコストが高くても放送波を用いて更新伝播を行うことにし、時間的に余裕がある場合にはコストを抑えて通信回線を用いて更新伝播を行うという選択を行うことができる。

【０１６６】また、本発明の分散型データベースシステムの一貫性管理方法（請求項６）によれば、請求項１～５のいずれか一つに記載の分散型データベースシステムの一貫性管理方法において、判定工程が、一貫性管理のレベルを判定することに加えて、データの新規登録か否

かを判定し、新規登録であると判定した場合に、設定されているレベルに関係なく、第３レベルと判定するため、データを新規登録する際の処理効率を向上させることができる。

【０１６７】さらに、本発明のコンピュータ読み取り可能な記録媒体（請求項７）によれば、請求項１～６のいずれか一つに記載の分散型データベースシステムの一貫性管理方法の各工程をコンピュータに実行させるためのプログラムを記録したため、このプログラムをコンピュータに実行させることにより、分散型データベースシステムの一貫性管理のレベルを複数段階に分け、トランザクションの種類に応じて異なる一貫性管理を行うことができ、分散型データベースシステムの利便性の向上を図ることができる。

【図面の簡単な説明】

【図１】集中管理または分散管理と、強い一貫性管理または弱い一貫性管理との組み合わせによって、分散型データベースシステムを構築した場合の長所および短所が変化することを示した説明図である。

【図２】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。

【図３】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、（ａ）は一貫性管理のレベルを設定する手順を、（ｂ）は（ａ）に示すようにして設定した一貫性管理のレベルに基づいて一貫性管理方法を実行する手順を示すフローチャートである。

【図４】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、各サイトから主サイトに送信されるＲＥＱＵＥＳＴメッセージの説明図である。

【図５】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、主サイトから各サイトに送信されるコミットメッセージの説明図である。

【図６】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、主サイトから各サイトに送信されるアボートメッセージの説明図である。

【図７】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、（ａ）は差異限定制約条件を設定する手順を、（ｂ）は（ａ）に示すようにして設定した差異限定制約条件に基づいて更新伝播を行う手順を示すフローチャートである。

【図８】本発明の実施の形態１に係る分散型データベースシステムの一貫性管理方法において、一貫性管理のレベルＡ～Ｃの特徴をまとめた説明図である。

【図９】本発明の実施の形態２に係る分散型データベースシステムの一貫性管理方法において、（ａ）は一貫性管理のレベルおよび新規登録トランザクションを設定する手順を、（ｂ）は（ａ）に示すようにして設定した一貫性管理のレベルおよび新規登録トランザクションの設定に基づいて一貫性管理方法および新規登録を実行する手順を示すフローチャートである。

【図１０】本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法を実現するための分散型データベースシステムの概念構成図である。

【図１１】本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法において、更新伝播タイミング設定処理を示すフローチャートである。

【図１２】本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法において、更新伝播タイミングをサイト側で設定する場合に、リクエストメッセージに添付されるデータを示す説明図である。

【図１３】本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法において、主サイトによる更新伝播管理処理を示すフローチャートである。

【図１４】本発明の実施の形態３に係る分散型データベースシステムの一貫性管理方法において、主サイトによって管理される更新伝播管理情報の一例を示す説明図である。

【符号の説明】

ＤＢＳ，　　主サイト

ＤＢＳ$_1$，ＤＢＳ$_2$，ＤＢＳ$_3$，・・・，ＤＢＳ$_n$　　サイト

１０００　　衛星

１００１ａ～１００１ｅ　　アンテナ

１００２　　通信回線

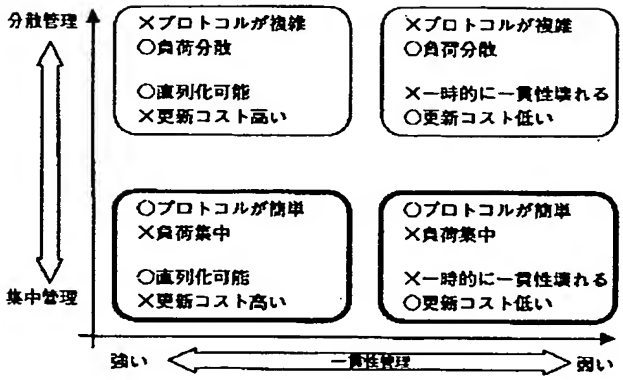
【図４】

REQUEST
・サイトID
・トランザクションID
・（操作、対象オブジェクトID、ｔ（O_i、T）【更新後の値】）の3組または４組のリスト

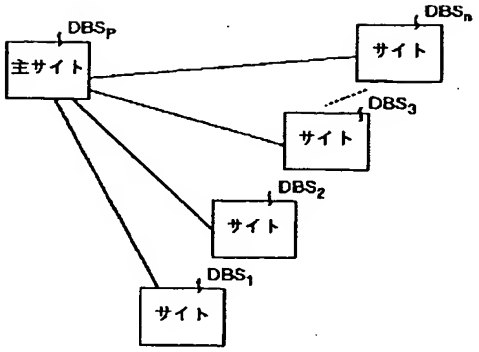
【図５】

COMMIT
・サイトID
・トランザクションID
・（対象オブジェクトID、ｔ（O_o）、更新後の値）の３組のリスト

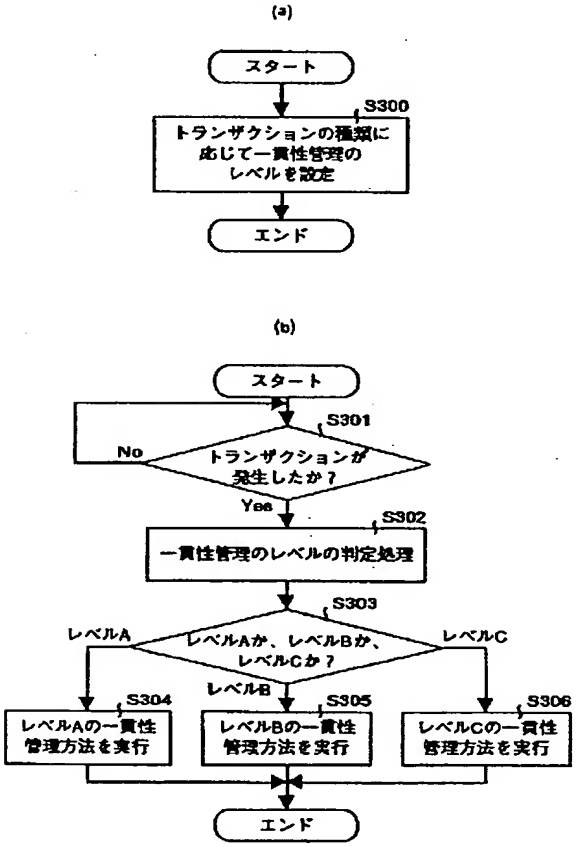
【図１】



【図２】



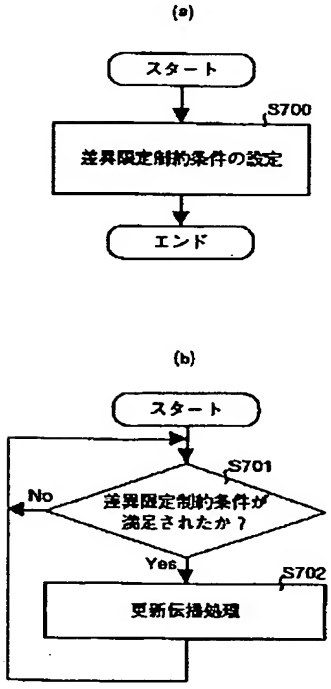
【図３】



【図６】

ABORT
・サイトID
・トランザクションID
・（対象オブジェクトID、t（O_o）、現在の値O_o）の３組のリスト

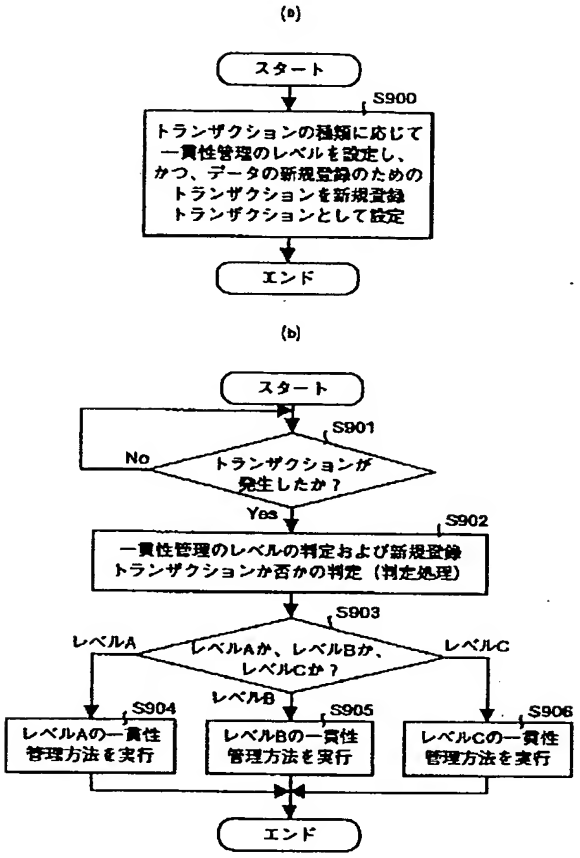
【図７】



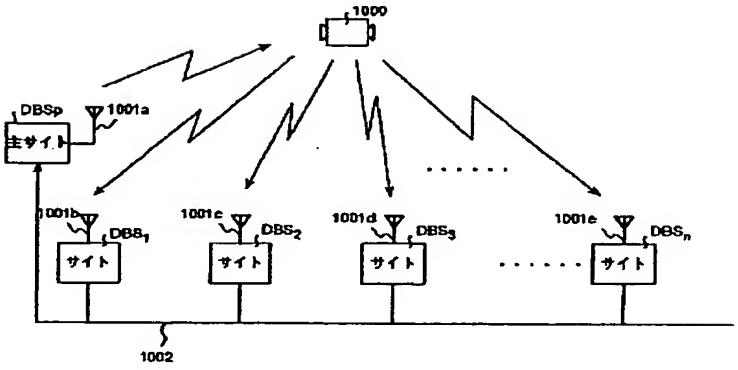
【図８】

レベル	read-onlyの トランザクション	更新を含む トランザクション	主サイトからの 更新伝播の遅延
A	厳格	厳格	即時
B	緩和 １コピー直列化可能	厳格	遅延させる （差異限定制約）
C	緩和 １コピー直列化可能	緩和	遅延させる （差異限定制約）

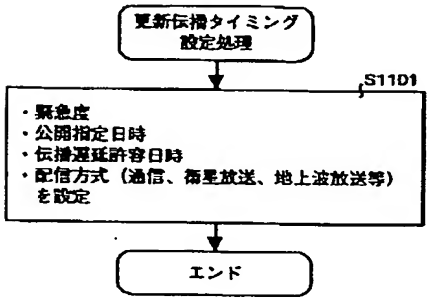
【図９】



【図１０】



【図１１】



【図12】

(a)

依頼元ID	送付先（主サイト）ID	依頼日時	更新OIDリスト	緊急度	公開指定日時	伝播遅延許容日時	配信方式
8945	56	1998/3/10 20:18:10	45,682,589,439,902	1	指定無し	1998/3/10 20:30:00	衛星

(b)

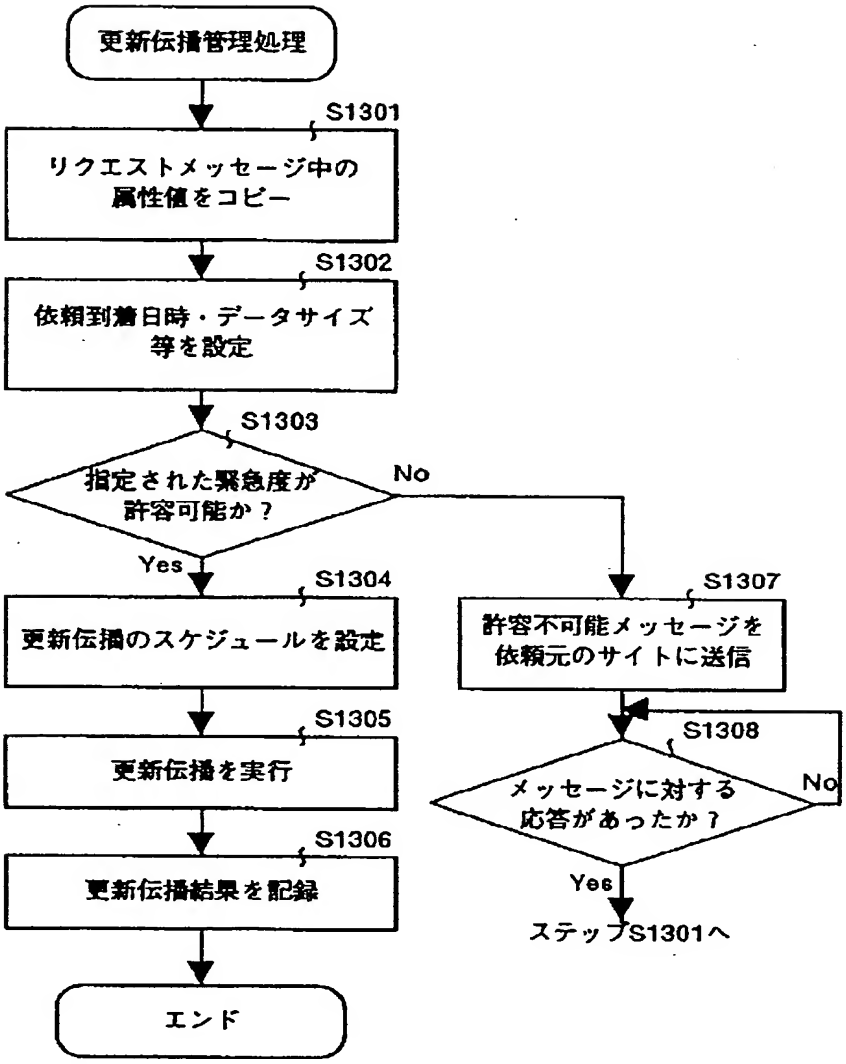
依頼元ID	送付先（主サイト）ID	依頼日時	更新OIDリスト	緊急度	公開指定日時	伝播遅延許容日時	配信方式
9824	90	1998/3/10 20:23:15	892,678	1	1998/4/1 10:00:00	1998/4/1 10:10:00	通信

【図14】

依頼元ID	依頼到着日時	依頼日時	更新OIDリスト	緊急度	公開指定日時	伝播遅延許容日時	データサイズ
8945	1998/3/10 20:19:13	1998/3/10 20:18:10	45,682,589,439,902	1	指定無し	1998/3/10 20:30:00	310MB

送付サイト数	配信方式	伝播開始日時	伝播完了日時	配信の遅れ
120	衛星	1998/3/10 20:25:21	1998/3/1020:25:21	0

【図１３】



フロントページの続き

(51) Int.Cl.[6]	識別記号	FI
		G06F 15/401　　340A

(72)発明者　前田　薫
東京都大田区中馬込１丁目３番６号　株式会社リコー内

(72)発明者　池田　哲也
東京都大田区中馬込１丁目３番６号　株式会社リコー内